SERVIR West Africa Training Program on Data Stewardship and the CoreTrustSeal Requirements: Session 1

Instructor: Robert R. Downs, PhD

Center for International Earth Science Information Network (CIESIN)

The Earth Institute, Columbia University

Coordinated by AFRIGIST

May 5, 2021, 10:00 a.m. EDT - 11:30 a.m. EDT (14:00 - 15:30 UTC)

Opening and Introduction to SERVIR West Africa Training Program on Data Stewardship and CoreTrustSeal Requirements

Schedule of Training Program Sessions

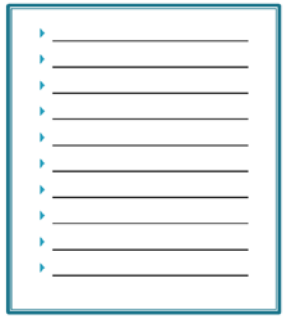
- Dates: May 5 (Session 1), May 12 (Session 2), May 26 (Session 3)
- Time: 10:00 a.m. to 11:30 a.m. EDT (14:00 - 15:30 UTC)
- Zoom Link: https://columbiauniversity.zoom.us/j/94522663326?pwd=M3FxL2ZOUTE2K1lUcjA4ZHdhdGNqdz09
- Meeting ID: 945 2266 3326 Passcode: 231758

Instructor

- Dr. Robert R. Downs
- Senior Digital Archivist at CIESIN
  - Center for International Earth Science Information Network, Earth Institute, Columbia University
- https://www.earth.columbia.edu/users/profile/robert-r-downs

Opening and Introduction to Training Program

- Key topics:
  - The approach to institutional self-assessment of data stewardship
  - Data product and service management
  - Planning for CoreTrustSeal certification
- Instructional format
  - Lecture, questions, and discussion
  - Homework assignments
- Sessions will be recorded
  - Recordings will be posted on the SERVIR West Africa website



Outline for Current Session

- Opening and Introduction
- Data Stewardship Concepts, Principles, and Certification Instruments
- Introduction to the CoreTrustSeal Certification requirements
- CoreTrustSeal Requirements R0, R1, R3, R5 and R6
- CoreTrustSeal Requirements R2 and R4
- Assignment: Describe How Data Repository Meets CoreTrustSeal Requirements R1, R2, R5, and R6

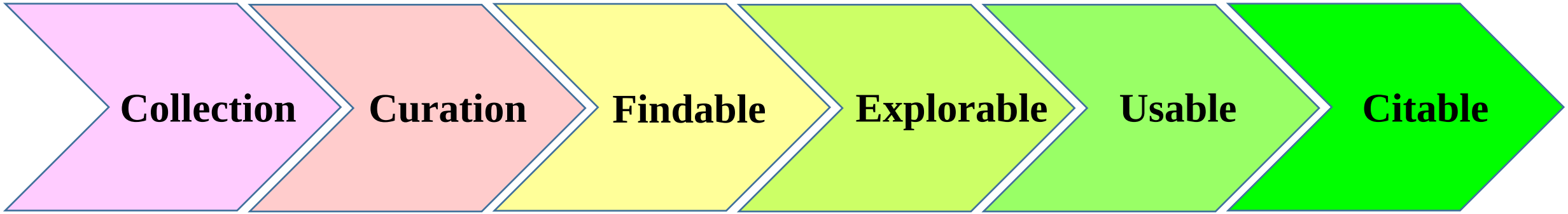
Data Stewardship Concepts, Principles, and Certification Instruments: Data Value

- Data are important and valuable information resources
  - Observations recorded by investing knowledge, time, and resources
  - Often contain unique observations from a specific time and place
- Value is realized when data are used
  - Conducting various kinds of studies, today, and in the future
  - Facilitating decision making and planning
  - Education and life-long learning
  - Practical applications by the general public

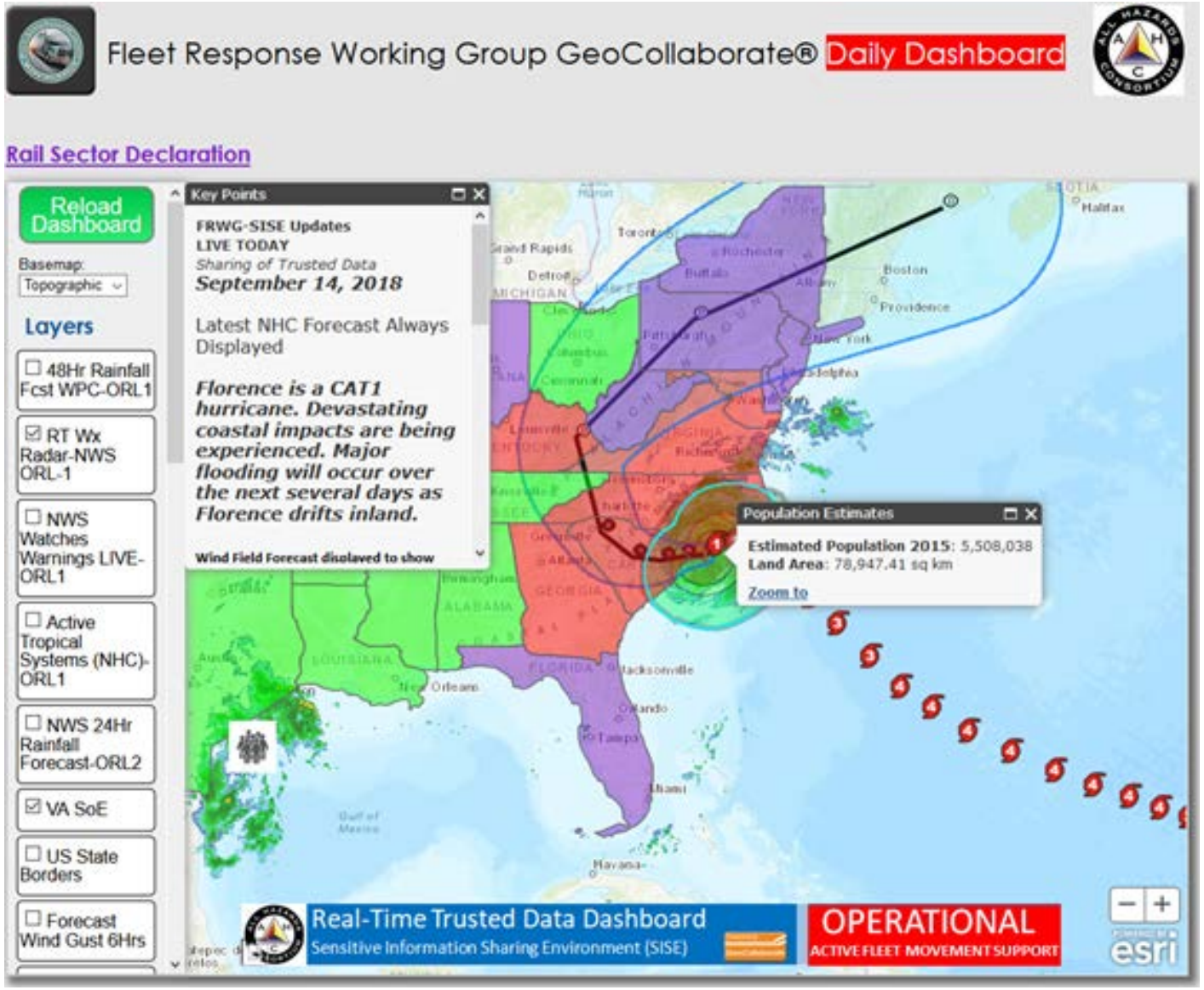
Data Stewardship Concepts, Principles, and Certification Instruments: Open Data

- Data value increases when data are open
  - Available for anyone to use for any purpose, without restrictions
  - Marked as open data with the terms of use clearly described
- Benefits of open data - reuse
  - Reuse by others, beyond the initial use by data collectors
  - Return on investment (ROI) increases with data reuse
  - Scientific value: new studies (hypotheses, questions)
  - Societal value, leading to improved
    - Healthcare
    - Communications
    - Transportation
    - Quality of life
    - …

Facilitating Data Value



Use Case on the Value of Reusing Open Data: Disaster Recovery

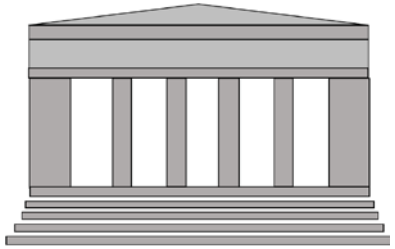


Collaboratively integrating NASA population data in a tool that enables utility crews to identify populations at risk of a hurricane

Data services enable disaster planners and responders to answer questions

- How many people will be affected by a flood?
- How many people live in close proximity to a nuclear power plant?

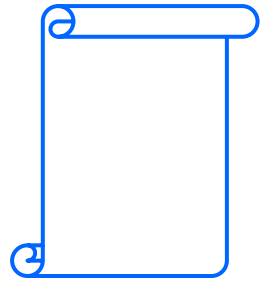
Data Stewardship Concepts, Principles, and Certification Instruments: Data Repositories



- Data centers, archives, museums, or other facilities
  - Domain repositories serve a particular discipline or subject area
  - Institutional repositories serve an organization or cooperating organizations
- Infrastructure for data stewardship
  - Archiving - Ingest, inspection, validation, packaging, preservation
  - Curation - Preparing data products and services for current and future use
  - Dissemination - Enabling discovery, exploration, access, and use
  - Support - Assisting users in accessing and using data and related resources

Data Stewardship Concepts, Principles, and Certification Instruments: Open Archival Information System (OAIS) Reference Model

- Framework for responsible stewardship of information resources
  - Defines functional requirements, terminology, and relationships between functional entities
  - Describes preservation perspectives and archive interoperability
- Developed by Consultative Committee for Space Data Systems (CCSDS)
  - Published by International Standard Organization (ISO) in 2003, revised in 2012
  - Currently being reviewed for revision
- ISO 14721:2012
  - Space data and information transfer systems — Open archival information system (OAIS) — Reference model
- Freely available from CCSDS
  - http://public.ccsds.org/publications/archive/650x0m2.pdf



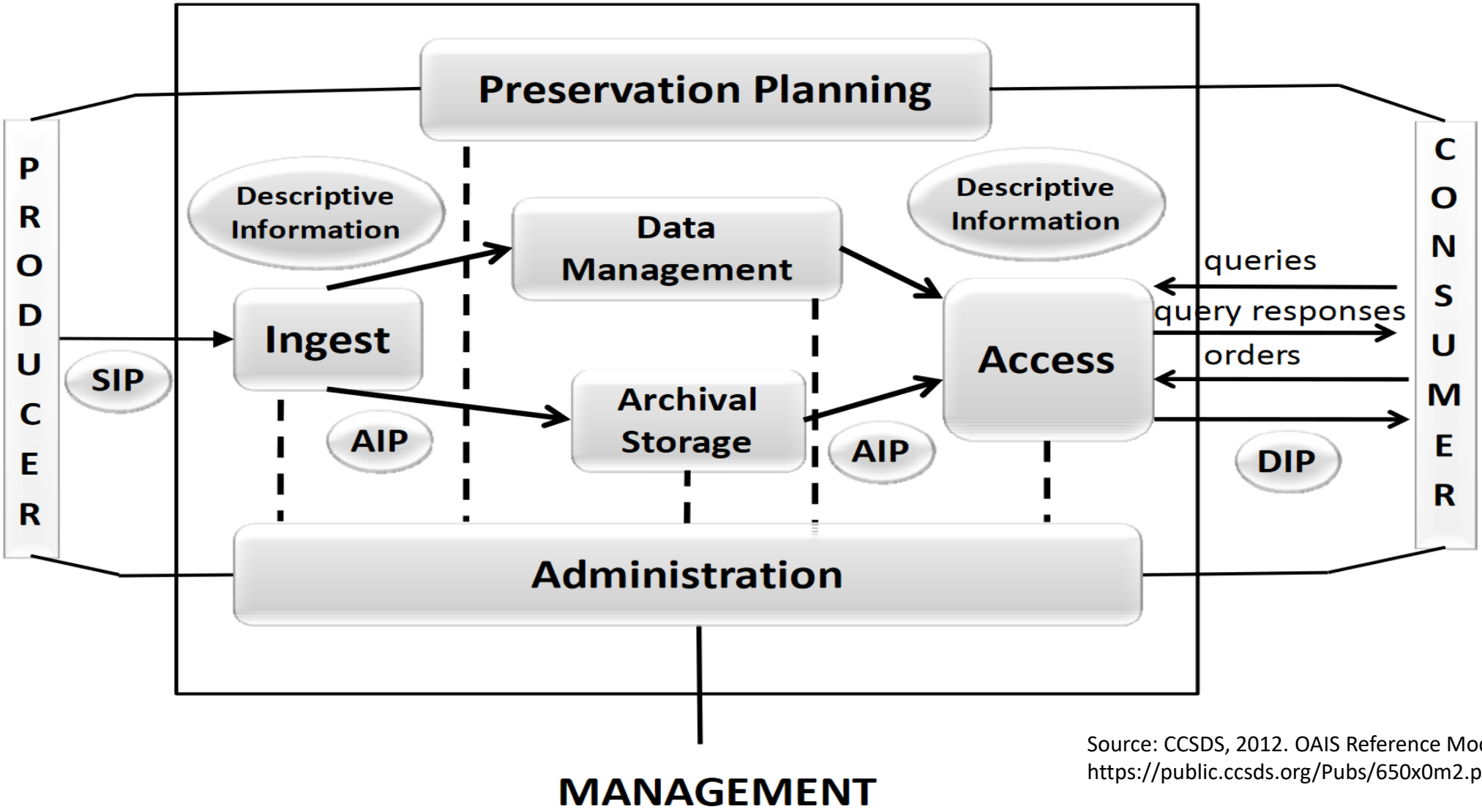
Data Stewardship Concepts, Principles, and Certification Instruments: Open Archival Information System (OAIS) Functional Requirements

- **Negotiates for and Accepts Information**
  - Negotiate for and accept appropriate information from information Producers.
- **Obtains Sufficient Control**
  - Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation.
- **Determines Designated Community**
  - Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base.
- **Ensures Information is Independently Understandable**
  - Ensure that the information to be preserved is Independently Understandable to the Designated Community. ... Designated Community should be able to understand the information without needing special resources … .
- **Follows Established Preservation Policies and Procedures**
  - Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed …
- **Makes the Information Available**
  - Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity.

Source: CCSDS. 2012. Reference Model for an OAIS. https://public.ccsds.org/Pubs/650x0m2.pdf

Data Stewardship Concepts, Principles, and Certification Instruments: Open Archival Information System (OAIS) - Continued

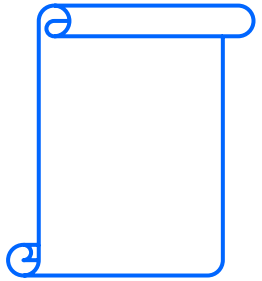
OAIS Functional Entities



Source: CCSDS, 2012. OAIS Reference Model. Figure 4-1. https://public.ccsds.org/Pubs/650x0m2.pdf

Data Stewardship Concepts, Principles, and Certification Instruments: Audit and Certification of Trustworthy Digital Repositories

- 125 requirements (metrics) hierarchically organized into 3 sections
  - Organizational Infrastructure, Digital Object Management, and Infrastructure and Security Risk Management
- Developed by the Consultative Committee for Space Data Systems
- Published as ISO 16363:2012. Space data and information transfer systems — Audit and certification of trustworthy digital repositories
- Currently under review by CCSDS and ISO
- http://public.ccsds.org/publications/archive/652x0m1.pdf



Data Stewardship Concepts, Principles, and Certification Instruments: TRUST Principles for Digital Repositories

- Transparency
  - To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
- Responsibility
  - To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
- User Focus
  - To ensure that the data management norms and expectations of target user communities are met.
- Sustainability
  - To sustain services and preserve data holdings for the long-term.
- Technology
  - To provide infrastructure and capabilities to support secure, persistent, and reliable services.

Source: Lin, et al. 2020. The TRUST Principles for digital repositories. https://doi.org/10.1038/s41597-020-0486-7

Data Stewardship Concepts, Principles, and Certification Instruments: FAIR Principles

- Describe characteristics of data and metadata to facilitate their use
  - Focus on data and offer guidance for data stewardship

Findable Accessible Interoperable Reusable

- Findable
  - Can the data be discovered within catalogs and the repository?
- Accessible
  - Can potential users render the data using a service or common tool?
- Interoperable
  - Can the data be used and integrated with other data products or services?
- Reusable
  - Can someone from outside the data development team use the data?

Based on Wilkinson, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. https://doi.org/10.1038/sdata.2016.18

Data Stewardship Concepts, Principles, and Certification Instruments: CARE Principles for Indigenous Data Governance

- Data created by or about Indigenous Peoples and their territories
  - Focus on data with implications for data stewardship and secondary use

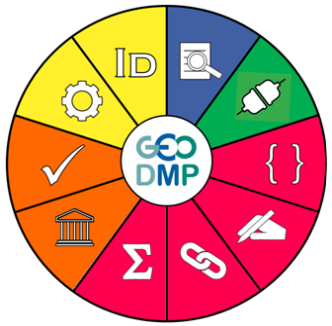
- **C**ollective Benefit
  - Inclusive development, improved governance, and equitable outcomes
- **A**uthority to **C**ontrol
  - Recognizing rights, interests and governance of data
- **R**esponsibility
  - Positive relationships, languages, increasing opportunities,
- **E**thics
  - Reducing potential harm, facilitating benefits, justice and future use

Based on Carroll, et al. 2020. The CARE Principles for Indigenous Data Governance. http://doi.org/10.5334/dsj-2020-043

Data Stewardship Concepts, Principles, and Certification Instruments: Group on Earth Observations (GEO)

Global Earth Observing System of Systems (GEOSS) Principles

- GEOSS Data Sharing Principles
  - Advocate data sharing practices to facilitate the use of open data
- GEOSS Data Management Principles
  - Advocate data management practices to facilitate data use and reuse



Source: GEO and Open EO Data. https://www.earthobservations.org/open_eo_data.php

Data Stewardship Concepts, Principles, and Certification Instruments: GEOSS Data Sharing Principles

GEOSS Data Sharing Principle 1.

Data, metadata and products will be shared as Open Data by default, by making them available as part of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE) without charge or restrictions on reuse, subject to the conditions of registration and attribution when the data are reused;

GEOSS Data Sharing Principle 2.

Where international instruments, national policies or legislation preclude the sharing of data as Open Data, data should be made available with minimal restrictions on use and at no more than the cost of reproduction and distribution;

GEOSS Data Sharing Principle 3.

All shared data, products and metadata will be made available with minimum time delay.

Derived from: GEOSS Data Sharing Principles Implementation Guidelines 2016-2025.

GEO Data Management Principles

Discovery

DMP-1: Metadata for Discovery

Accessibility

DMP-2: Online Access

Usability

DMP-3: Data Encoding

DMP-4: Data Documentation

DMP-5: Data Traceability

DMP-6: Data Quality-Control

Preservation

DMP-7: Data Preservation

DMP-8: Data and Metadata Verification

Curation

DMP-9: Data Review and Reprocessing

DMP-10: Persistent and Resolvable Identifiers

Derived from: Group on Earth Observations. 2016. Data Management Principles Implementation Guidelines.

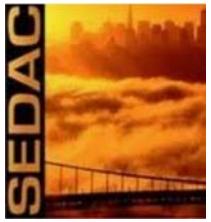
Enabling Data Discovery by Implementing DMP-1 Metadata for Discovery

Examples from the NASA SEDAC
Complete metadata for each dataset

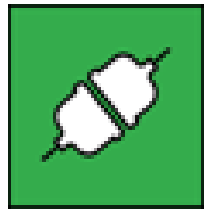
- Complete and correct metadata to enable discovery
  - Abstract and recommended citation on landing page and in metadata
- Metadata accessible from each dataset landing page
  - HTML, XML, Text
- Simplified language for metadata elements
  - E.g. Standardized rights declaration language, location names, keywords
- Metadata populating multiple catalogs
  - SEDAC, NASA Common Metadata Repository, DataCite

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Facilitating Data Access by Implementing DMP-2: Online Access

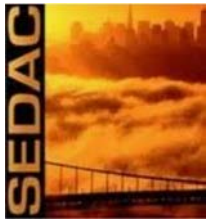


Examples from NASA SEDAC
24/7 Online Access to Open Data from SEDAC Website



- Negotiate permissions to enable access to open data
  - Applied CC By license to maps and GPWv4 collection of datasets
  - Attribution only required for many datasets
- Landing page for each data collection
  - Collection Overview, Datasets, Map Gallery, Map Services
  - Also distributing: Citations, FAQs, Acknowledgements
- Landing page for each dataset
  - Dataset Overview: Title, Purpose, Abstract, Recommended Citation, Available Formats
- Dataset landing page tabs
  - Dataset Overview, Download, Maps, Map Services, Documentation, Metadata

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.



Socioeconomic Data and Applications Center (SEDAC)

A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIESIN at Columbia University

Search SEDAC | Data | Login

DATA | MAPS | THEMES | RESOURCES | SOCIAL MEDIA | ABOUT | HELP

Gridded Population of the World (GPW), v4

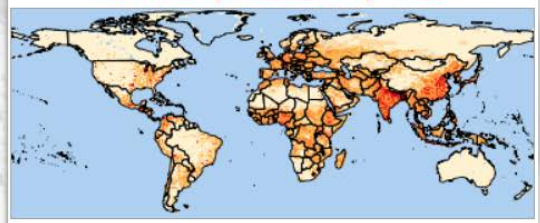
Follow Us: | Share:

- Collection Overview
- Methods
- Data Sets (8)
  - *Population Density, v4 (2000, 2005, 2010, 2015, 2020)*
  - Show All...
- Map Gallery (27)
- Map Services (26)
- Citations
- FAQs
- What's New in GPWv4?
- Documentation
- What is UN-adjusted population data?
- Multimedia
- Acknowledgments
- SEDAC Hazards Mapper
- Population Estimation Service

Population Density, v4 (2000, 2005, 2010, 2015, 2020)

Set Overview | Data Download | Maps | Map Services | Documentation | Metadata

GPWv4: Population Density - 2020



5 of 5

Purpose:

To provide estimates of population density for the years 2000, 2005, 2010, 2015, and 2020, based on counts consistent with national censuses and population registers, as raster data to facilitate data integration.

Abstract:

Gridded Population of the World, Version 4 (GPWv4) Population Density consists of estimates of human population density based on counts consistent with national censuses and population registers, for the years 2000, 2005, 2010, 2015, and 2020. A proportional allocation gridding algorithm, utilizing approximately 12.5 million national and sub-national administrative units, is used to assign population values to 30 arc-second (~1 km) grid cells. The population density grids are created by dividing the population count grids by the land area grids. The pixel values represent persons per square kilometer.

Recommended Citation(s)*:

Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Density. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4NP22DQ. Accessed DAY MONTH YEAR.

ENW (EndNote & RefWorks)†
RIS (Others)

** When authors make use of data they should cite both the data set and the scientific publication, if available. Such a practice gives credit to data set producers and advances principles of transparency and reproducibility. Please visit the data citations page for details. Users who would like to choose to format the citation(s) for this dataset using a myriad of*

Fostering Data Usability by Implementing DMP-3: Data Encoding

Examples from the NASA SEDAC

Diverse Community Needs Drive Data Products and Services

- SEDAC data used across multiple disciplines
  - 22 Major Fields, including: Environment/ecology, social sciences, geosciences, biology & chemistry, economics & business, engineering, plant & animal science, multidisciplinary, clinical medicine
- Dissemination requirements from community needs
  - Literature, events, User Working Group, NASA and working groups
- Selection of data encodings
  - Data products: identification, selection, and review
  - Data formats based on community practice
  - Multiple formats disseminated for many datasets
- Development of tools and services
  - Evolve to reflect changes in community practice

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Fostering Data Usability by Implementing DMP-4: Data Documentation

Examples from the NASA SEDAC Data Documentation Template

Documentation for <Dataset Title>
<Documentation Publication Date>
<Authors>
Abstract
Data set citation
Suggested citation for documentation
Contact to provide feedback on documentation
Table of Contents

I. Introduction
II. Data and Methodology
III. Data Set Description(s)
IV. How to Use the Data
IV. How to Use the Data
V. Potential Use Cases
VI. Limitations
VII. Acknowledgments
VIII. Disclaimer
IX. Use Constraints
X. Recommended Citation(s)
XI. Source Code
XII. References
XIII. Documentation Copyright & License
Appendix 1. Contributing Authors & Documentation Revision History
Appendix 2. Data Revision History

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Fostering Data Usability by Implementing DMP-5: Data Traceability

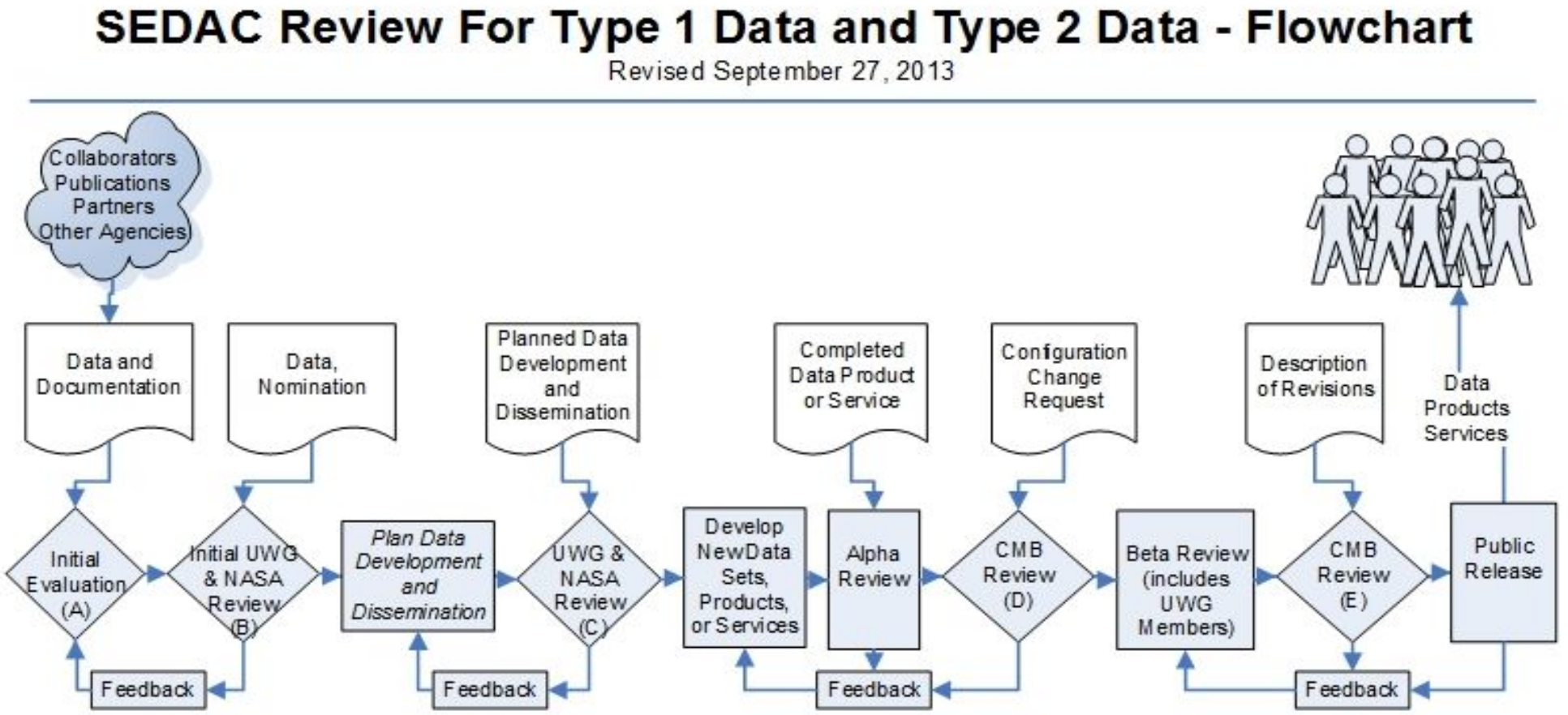
Examples from the NASA SEDAC
Capturing provenance of each dataset

- History of dataset creation and production
  - Permissions received
  - Data producer contact information
  - Communications with data producer
- Methodology described in documentation
  - Research design and original article referenced, if applicable
  - Data collection, derivation, and processing techniques
  - Instruments and software utilized
- Data revision history in documentation
  - Changes in each version
  - Errors corrected
  - Format conversions

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

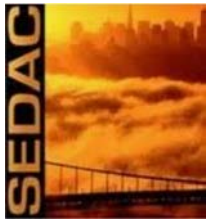
Fostering Data Usability by Implementing DMP-6: Data Quality-Control

Examples from the NASA SEDAC
Internal and External Reviews throughout Data Lifecycle



Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Fostering Data Preservation by Implementing DMP-7: Data Preservation



Examples from the NASA SEDAC Archiving and Assessments for Security and Sustainability

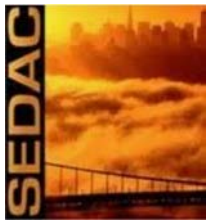
- Procedures for routinely accessioning each dataset
  - Appraisal, message digests, file & format inventories, redundant storage, etc.
- Selecting superseded data for SEDAC Long-Term Archive
  - Collaboratively established LTA with CU Libraries and Earth Institute
- Regular audits of data center
  - Routine NASA security and risk audits
  - Self-assessments and external test audit for ISO 16363 compliance
  - World Data System Certification

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Fostering Data Preservation by Implementing DMP-8: Data and Metadata Verification

Examples from the NASA SEDAC

Reviews of Content, Media, Platforms, Systems, and Hardware



- Procedures for verifying media integrity
  - Periodic inspections of access to content
  - Refreshment and testing
- Integrity of Archival Information Packages
  - Integrity verification with SHA-1 message digests
  - Changes to a dataset require creation of a new version
- Periodic review of hardware and system software
  - Replacement of components prone to failure
  - Market studies on media, servers, and platforms
- Preventative measures for security
  - Two factor authentication
  - Onsite security audits

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

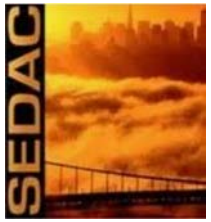
Fostering Data Curation by Implementing DMP-9: Data Review and Reprocessing

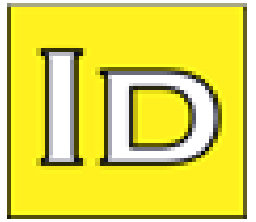
Examples from the NASA SEDAC Review Processes and Procedures

- Review procedures from pre-acquisition to production
  - Data Producers, Project Scientists, User Working Group, NASA, Beta Reviewers, Configuration Management Board
- Continuing post-production review process
  - Reference services (Help Desk) to receive questions and feedback
  - FAQ populated with answers to common questions
  - Recommendations reviewed for possible improvements
  - Creating new versions of data containing corrections

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Fostering Data Curation by Implementing DMP-10: Persistent and Resolvable Identifiers





Examples from the NASA SEDAC Digital Object Identifiers (DOI) for Persistent Identifiers

- Routinely assigning a DOI for each dataset
  - Policy guidance and procedures for assigning and maintaining DOIs
- DOIs of Related Resources in DataCite Metadata
  - Linking data to documentation, related data products, and publications
- Assigning a DOI to new documentation
  - Exploring additional opportunities for persistent identifiers

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Practical Approaches for Repositories to Improve Data Management

- Develop a template to guide the documentation of data products and begin documenting data early in the data lifecycle
- Involve data users and other stakeholders in reviews of data products at each stage throughout the data lifecycle
- Assign a persistent identifier to each data product and include it in recommended data citation that displays on the data landing page
- Conduct self-assessments and external audits using instruments that measure capabilities of trustworthy repositories

Source: Downs, 2017. Implementing the Group on Earth Observations Data Management Principles: Lessons from a Scientific Data Center. ISPRS 2017.

Example: SEDAC Data Documentation Template

Documentation for <Dataset Title>
<Documentation Publication Date>
<Authors>
Abstract
Data set citation
Suggested citation for documentation
Contact to provide feedback on documentation
Table of Contents

Why Audit Research Data Centers?

- **Data producers** need to know where to deposit their data
  - Trust that their data will be preserved, curated, and disseminated
- **Data users** need to know where they can find data
  - Data that are vetted, described for use, and available in the future
- **Funders** need to know who to support for data management
  - Where services are reviewed routinely for continuous improvement
- **Publishers** need to know who to recommend for archiving data
  - Where referenced data will be persistently accessible and usable
- **Data professionals** need to know where they can practice
  - Apply their data management skills
  - Obtain professional development in data stewardship
- **Data centers** need to know how they are performing
  - Policies, procedures, and practices that need improvement

Source: Downs, 2016. Audit of a Scientific Data Center: Opportunities and Approaches. SciDataCon 2016.

Data Repository Audit: Cost vs Benefits

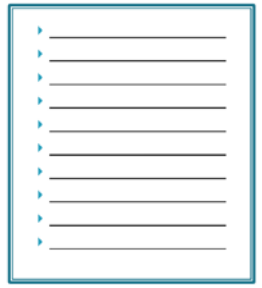
- Potential costs
  - **Time** for repository management and staff training
  - **Training** registration fees and travel costs, if applicable
  - **Time** for repository management and staff to prepare and be audited
  - Pre-audit improvements (**staff, software, hardware**), if applicable
  - Audit instrument **purchase fee**, if applicable
  - Audit **service fee**, including auditor travel costs, if applicable

- Potential benefits
  - Improve **transparency** and **quality assurance**
  - **Compare capabilities and services** with standard practices
  - **Review and identify gaps** in current and planned services
  - **Plan for needed enhancements** to improve capabilities and services
  - Obtain **certification** that recognizes attainment of standard, if applicable

Source: Downs, 2016. Audit of a Scientific Data Center: Opportunities and Approaches. SciDataCon 2016.

Approaching the Repository Audit

- Planning
  - Initiated by repository management
  - Identify candidate audit instruments
- Preparation
  - Conduct a self-assessment
  - Complete improvements to address weaknesses
- Scheduling
  - Identify availability of auditors
  - Identify availability of repository management and key staff



Source: Downs, 2016. Audit of a Scientific Data Center: Opportunities and Approaches. SciDataCon 2016.

Selecting the Audit Instrument

- Measurement: what does the repository want to measure?
  - What is the scope of the planned assessment? data management, security, staff capabilities, efficiency, finances, sustainability, …
  - Are there instruments available to assess the area of interest?

Source: Downs, 2016. Audit of a Scientific Data Center: Opportunities and Approaches. SciDataCon 2016.

Evaluating Candidate Audit Instruments

- What does the candidate instrument measure?
  - **Is it applicable** to the data center, its services, and capabilities?
  - **Can the metrics measure** whether each requirement has been satisfied?
- Can the instrument be used internally?
  - **By data center staff** for pre-assessment or post-assessment reviews
- Validity
  - Has the instrument been **developed** by a reputable organization?
  - Has the instrument been **reviewed** and reviewed recently?
  - Is the instrument a **recognized stan**dard for assessing data centers?
  - Has it been **endorsed** by the community and by independent bodies?

Source: Downs, 2016. Audit of a Scientific Data Center: Opportunities and Approaches. SciDataCon 2016.

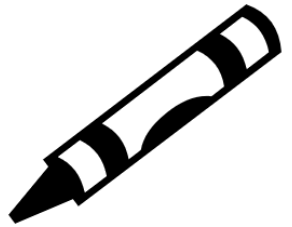
Data Repository Audit Considerations

Frequency	Authority: Internal	Authority: External
Once	Internal one-time self-assessment	External one-time audit
Periodic	Internal periodic self-assessments	External periodic audits

- Scope of audit:
  - Holistic vs targeted to specific capabilities, functions, or collections
- Approach should be based on objectives for the assessment
  - Why is the repository seeking an audit?
  - Which stakeholders are encouraging the audit?
  - Is improvement the primary objective or a credential?

Source: Downs, 2016. Audit of a Scientific Data Center: Opportunities and Approaches. SciDataCon 2016.

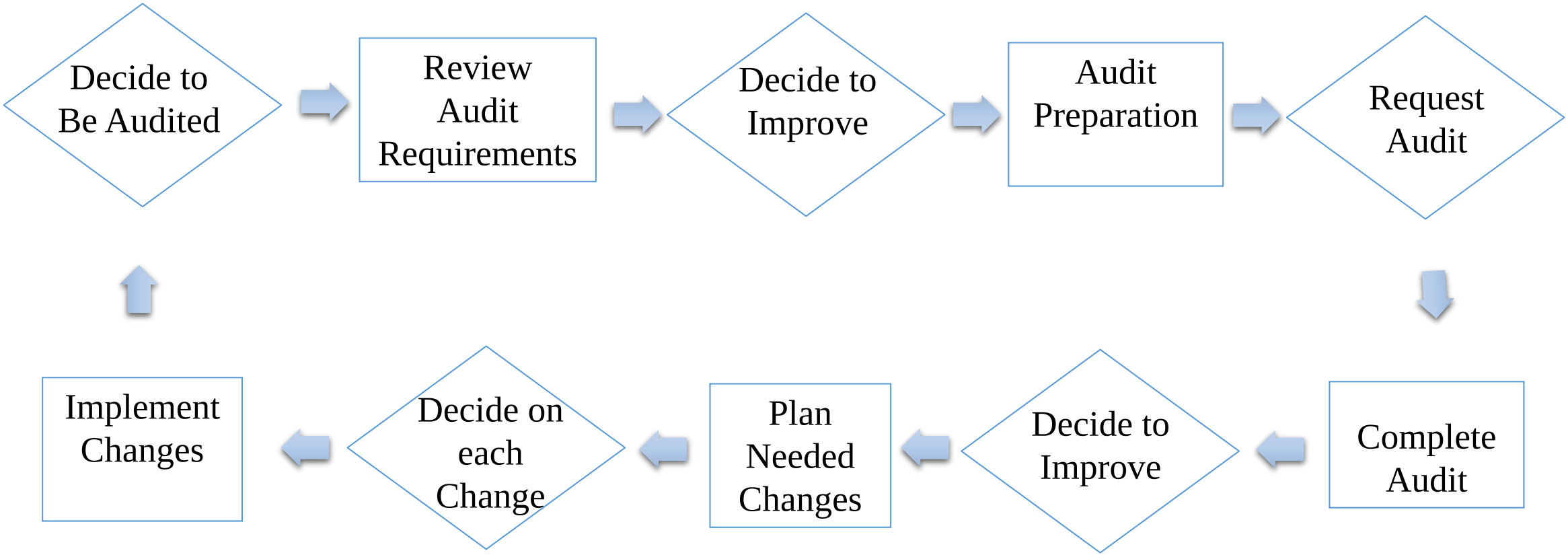
Data Repository Audit Opportunity



- Opportunity to identify, document, justify, and plan for needed enhancements, such as policy improvements, resource acquisitions, professional development, staff recruitment, procedure revisions, system upgrades, and new services

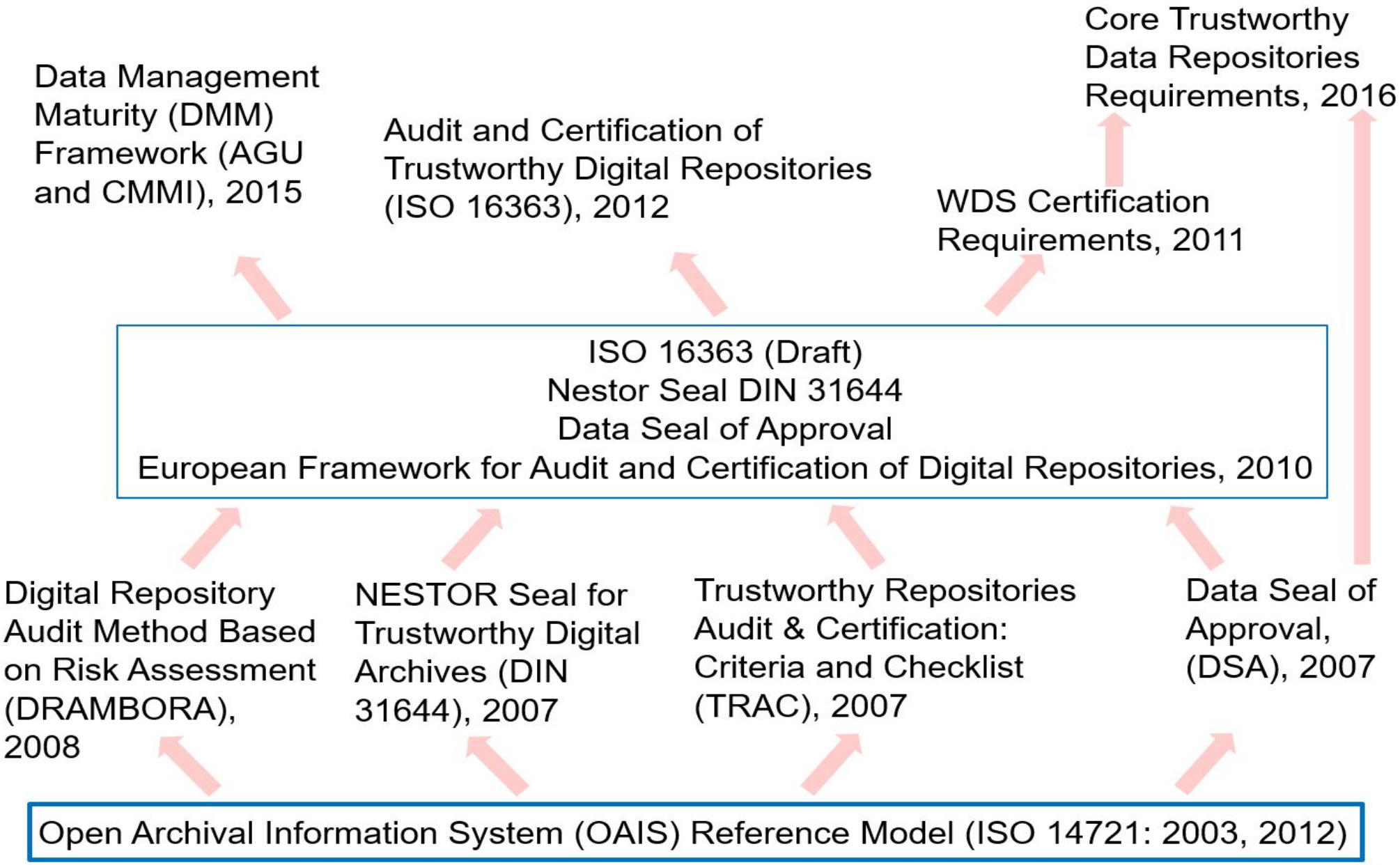
Source: Downs, 2016. Audit of a Scientific Data Center: Opportunities and Approaches. SciDataCon 2016.

Continuously Improving the Scientific Data Archive



Downs & Chen 2012 Improving the Trustworthiness of an Interdisciplinary Scientific Data Archive

Evolution of Instruments for Assessing Trustworthy Data Repositories



Source: Downs, 2019. International Standards for Trustworthy Data Repositories. https://doi.org/10.7916/d8-h7xf-ha70

Introduction to the CoreTrustSeal Certification Requirements: Using Requirements for Improvement, Self-Assessment and Certification

- Improvement
  - Identify opportunities for improving capabilities
- Self-Assessment
  - Identify areas to be improved and prepare for certification
- Certification
  - Apply for external review of repository capabilities and request certification
  - Validation period is 3 years from the date when certification is obtained

Introduction to the CoreTrustSeal Requirements: Compliance Levels

- 0 – Not applicable
- 1 – The repository has not considered this yet
- 2 – The repository has a theoretical concept
- 3 – The repository is in the implementation phase
- 4 – The guideline has been fully implemented in the repository

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

Introduction to the CoreTrustSeal Certification Requirements: Structure

- Background Information
  - R0 Context
- Organizational Infrastructure
  - R1 Mission/Scope, R2 Licenses, R3 Continuity of access, R4 Confidentiality/Ethics, R5 Organizational infrastructure, R6 Expert guidance
- Digital Object Management
  - R7 Data integrity and authenticity, R8 Appraisal, R9 Documented storage procedures, R10 Preservation plan, R11 Data quality, R12 Workflows, R13 Data discovery and identification, R14 Data Reuse
- Technology
  - R15 Technical Infrastructure, R16 Security

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirements R0, R1, R3, R5 and R6

- R0 Context
  - General information about the repository
- R1 Mission/Scope
  - Publicly accessible mission statement
- R3 Continuity of Access
  - Plan for continuing operations for the long-term and during unanticipated events
- R5 Organizational Infrastructure
  - Sufficient funding, staffing, governance, and structure to operate
- R6 Expert Guidance
  - Guidance received to ensure that the repository meets community needs

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirement R0 - Context

- Repository Type. Select all relevant types from:
  - Domain or subject-based repository; Institutional repository; National repository system, including governmental; Publication repository; Library; Museum; Archive; Research project repository; Other
- Brief Description of Repository
- Brief Description of the Designated Community
- Level of Curation Performed. Select all relevant types from:
  - Content distributed as deposited Basic curation – e.g., brief checking, addition of basic metadata or documentation Enhanced curation – e.g., conversion to new formats, enhancement of documentation; Data -level curation – as in C above, but with additional editing of deposited data for accuracy
- Insource/Outsource Partners. If applicable, please list them. –
- Summary of Significant Changes Since Last Application (if applicable) –
- Other Relevant Information

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirement: R1 Mission/Scope

- The repository has an explicit mission to provide access to and preserve data in its domain.
- Responses should include evidence related to the following questions:
  - Your organization's mission in preserving and providing access to data, and include links to explicit statements of this mission.
  - The level of approval that the mission has received within the organization.
- Evidence for this Requirement could take the form of
  - an *approved public mission statement*, *roles mandated by funders*, *policy statement signed off by governing board*.

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirement: R3 Continuity of Access

- The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.
- Responses should include evidence related to the following questions:
  - The level of responsibility undertaken for data holdings, including any guaranteed preservation periods.
  - The medium-term (three- to five-year) and long-term (> five years) plans in place to ensure the continued availability and accessibility of the data. In particular, both the response to rapid changes of circumstance and long-term planning should be described, indicating options for relocation or transition of the activity to another body or return of the data holdings to their owners (i.e., data producers). For example, what will happen in the case of cessation of funding, which could be through an unexpected withdrawal of funding, a planned ending of funding for a time-limited project repository, or a shift of host institution interests?

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirement: R5 Organizational Infrastructure

- The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission
- Responses should include evidence related to the following questions:
  - The repository is hosted by a recognized institution (ensuring long-term stability and sustainability) appropriate to its Designated Community.
  - The repository has sufficient funding, including staff resources, IT resources, and a budget for attending meetings when necessary. Ideally this should be for a three- to five-year period.
  - The repository ensures that its staff have access to ongoing training and professional development.
  - The range and depth of expertise of both the organization and its staff, including any relevant affiliations (e.g., national or international bodies), is appropriate to the mission.

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirement: R6 Expert Guidance

- The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).
- Responses should include evidence related to the following questions:
  - Does the repository have in-house advisers, or an external advisory committee that might be populated with technical, curation, data science, and disciplinary experts?
  - How does the repository communicate with the experts for advice?
  - How does the repository communicate with its Designated Community for feedback?

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

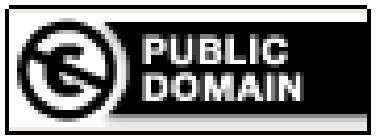
CoreTrustSeal Requirements R2 and R4

- R2 Licenses
  - Intellectual property rights and restrictions on data for repository and users
- R4 Confidentiality/Ethics
  - Compliance with disciplinary and community norms for confidentiality

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirement: R2 Licenses

- The repository maintains all applicable licenses covering data access and use and monitors compliance.
- Responses should include evidence related to the following questions:
  - License agreements in use.
  - Conditions of use (Intellectual Property Rights, distribution, intended use, protection of sensitive data, etc.).
  - Documentation on measures in the case of noncompliance with conditions of access and use.



Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

CoreTrustSeal Requirement: R4 Confidentiality and Ethics

- The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.
- Responses should include evidence related to the following questions:
  - How does the repository comply with applicable disciplinary norms?
  - Does the repository request confirmation that data collection or creation was carried out in accordance with legal and ethical criteria … (e.g., Ethical Review Committee …?
  - Are special procedures applied to manage data with disclosure risk?
  - Are data with disclosure risk distributed under appropriate conditions?
  - Are procedures in place to review disclosure risk … in a secure way?
  - Are staff trained in the management of data with disclosure risk?
  - Are there measures in place if conditions are not complied with?
  - Does the repository provide guidance in the responsible deposit, download, and use of disclosive, or potentially disclosive data?

Source: CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022. https://doi.org/10.5281/zenodo.3638211

Assignment: Describe How Data Repository Meets CoreTrustSeal Requirements R1, R2, R5, and R6

- Assess your data repository in terms of each requirement, focusing on the questions in the Extended Guidance
  - CoreTrustSeal Requirements: https://doi.org/10.5281/zenodo.3638211
  - Extended Guidance: https://doi.org/10.5281/zenodo.3632533
  - Glossary: https://doi.org/10.5281/zenodo.3632563
- Identify ways in which the repository meets each requirement
  - Prepare a short description of a capability or feature of the repository that meets one or more aspects of each CoreTrustSeal requirement
- Identify opportunities for improving the repository to meet each requirement
  - Prepare a short description of a feature that could be improved or implemented in the repository to address one or more aspects of each CoreTrustSeal requirement

The SERVIR 2021 Training Program on Data Stewardship and the CoreTrustSeal Requirements videos are available online on the NASA SEDAC YouTube Channel:

https://www.youtube.com/channel/UCjUjAvV7M04SxxpM5wq4fMw?view_as=public

Thank you!

rdowns@ciesin.columbia.edu